



# *Climate in a Box Workshop*

## *Workflow Tool*



NASA SIVO Code 610.3

September 22, 2010

**Presenters:**

March Andrews, Robert Burns, Megan Damon,  
John Evans, Shawn Freeman, Phil Hayes, Bruce  
Van Aartsen



## *Outline: Morning Session*

- Introduction to the NASA Workflow Tool
- NASA Experiment Designer (NED)
- Break
- Running the GEOS5 Workflow
- Creating a Workflow – Part I: Overview
- Lunch



## *Outline: Afternoon Session*

- Creating a Workflow – Part II: Hands-on Training
- Break
- Distributed Modeling System
- Workshop Summary



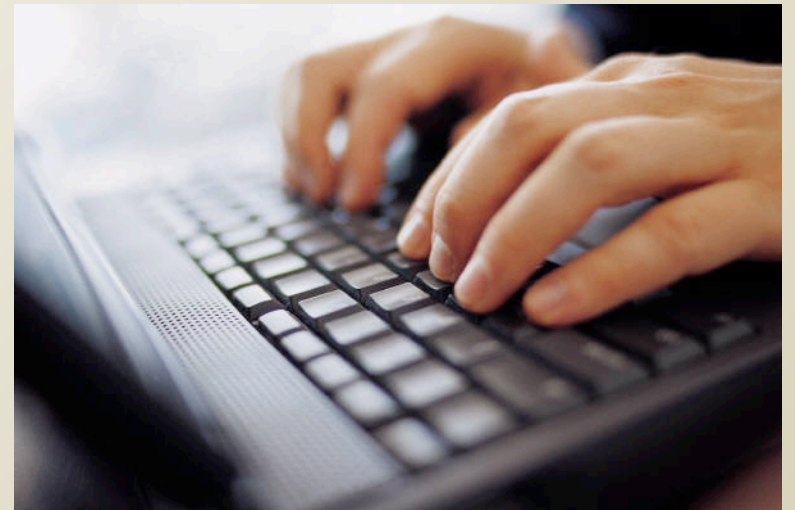
# *Introduction to the NASA Workflow Tool*



## *What is a Workflow Tool?*

A system that supports automating a series of tasks such as setting up and running complex model experiments.

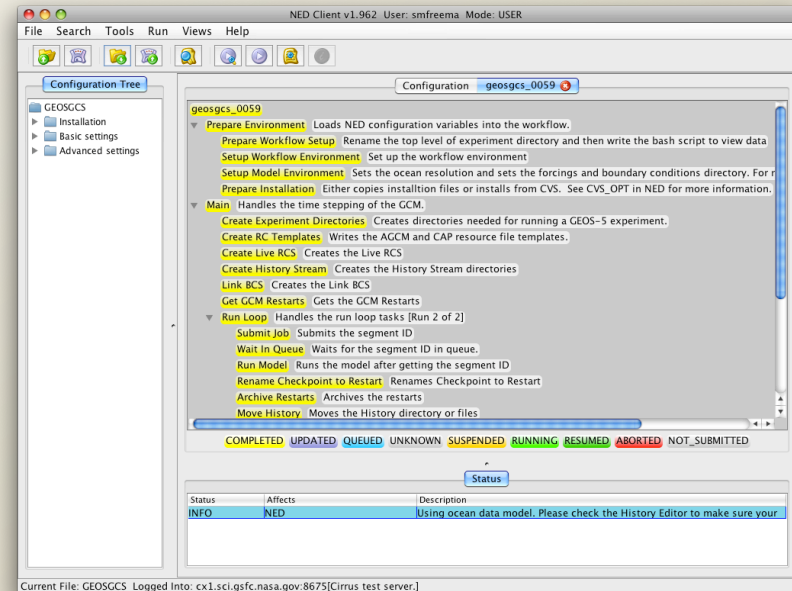
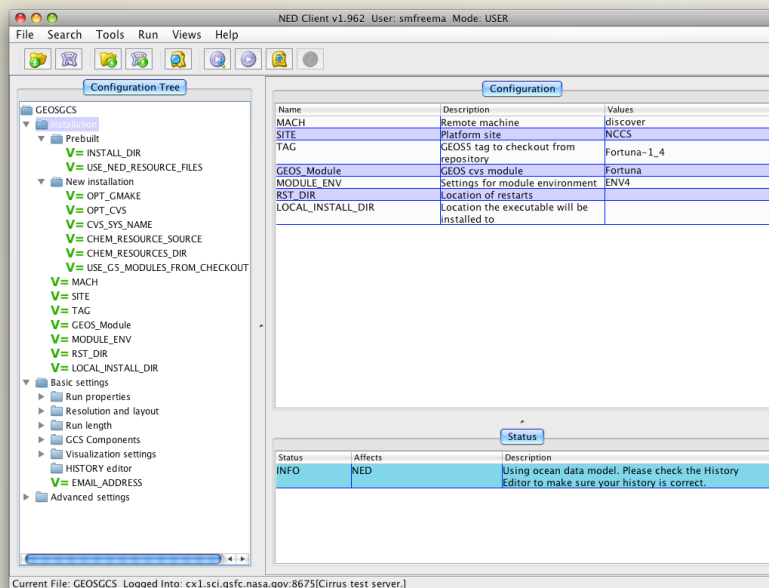
***Enable users to  
concentrate on the  
scientific results  
rather than the  
system  
environments.***

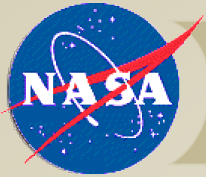




# NASA Workflow Tool: Easy Model Operation

- Configure, run and monitor models from a GUI
- Displays configuration and execution steps
- Removes need for user to have to manage: system environment, build settings, config files and other details
- Intended for using models -- not a replacement for highly customized modeling activities

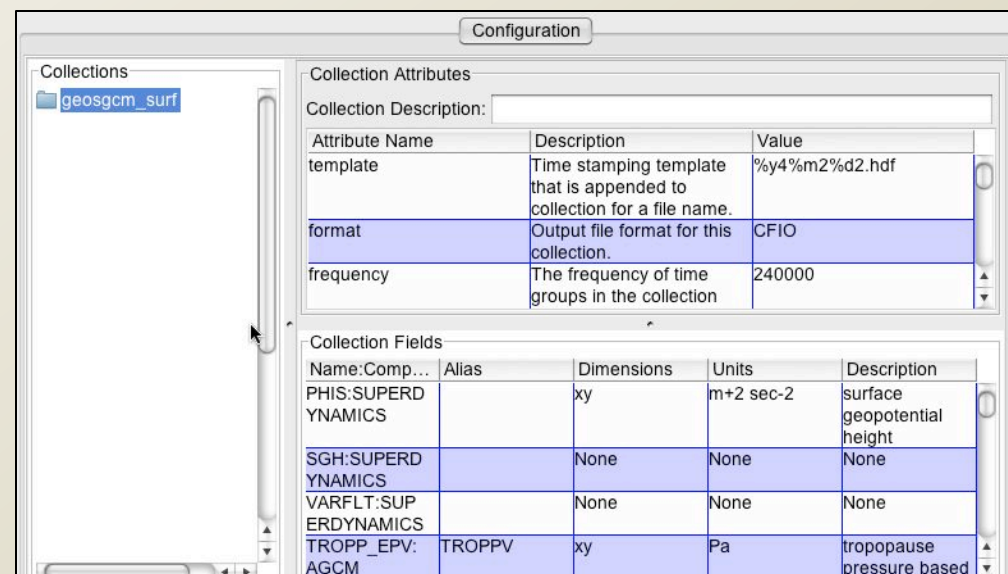
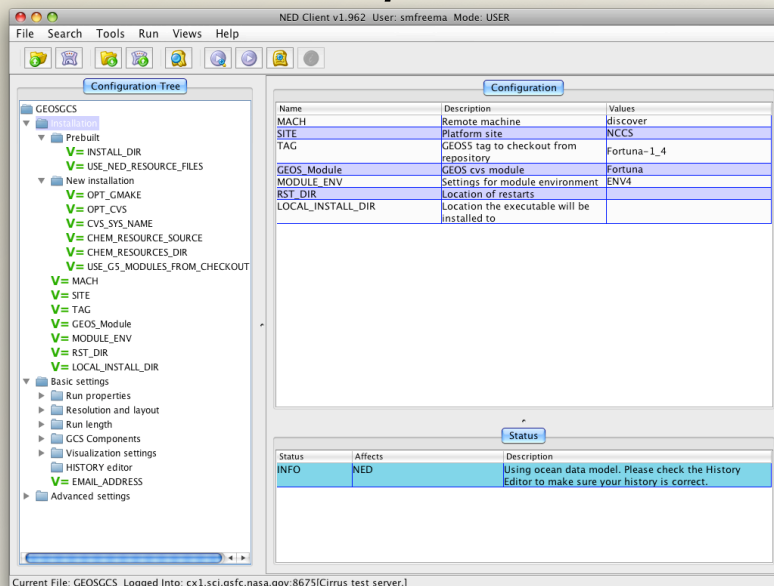




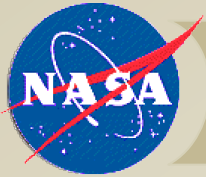
# NASA Workflow Tool:

## Configuration Assistance

- Model options are organized and presented for configuration and running the model
- Easier for non-specialists to run models
- Can script model configuration actions
  - *Example:* Detect configuration error before running



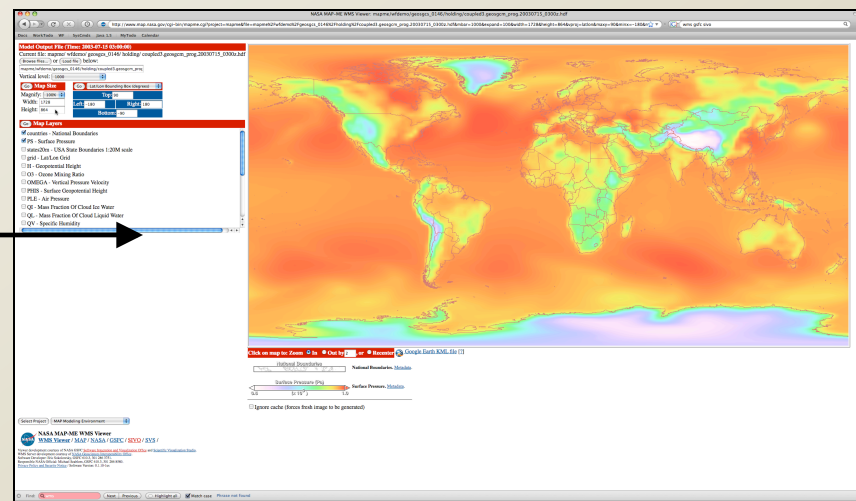
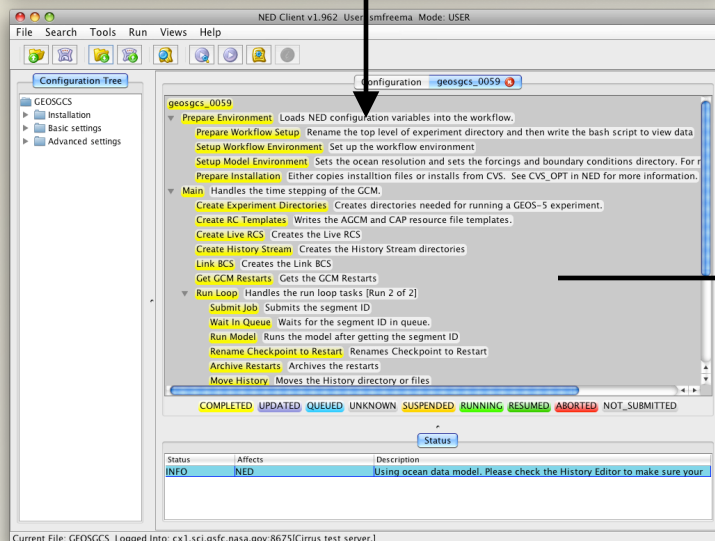
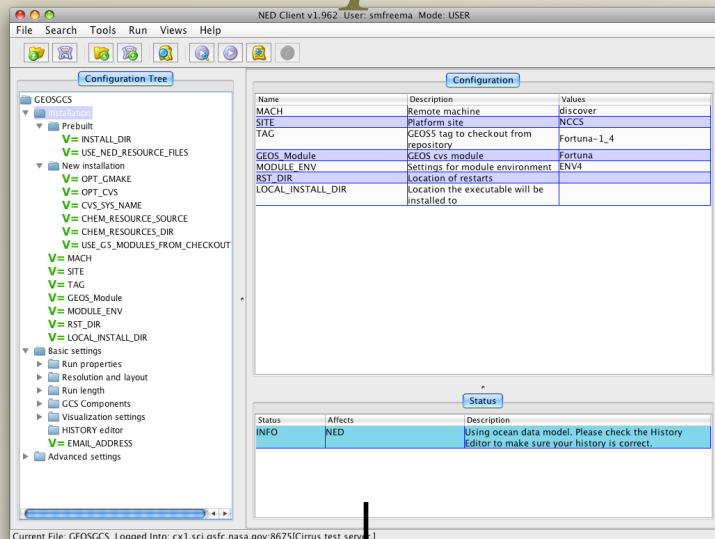




# NASA Workflow Tool:

## Experiment Management

- Save experiment configurations outputs for each run
- Load old experiments and rerun
- Allows experiments to be easily shared and repeated

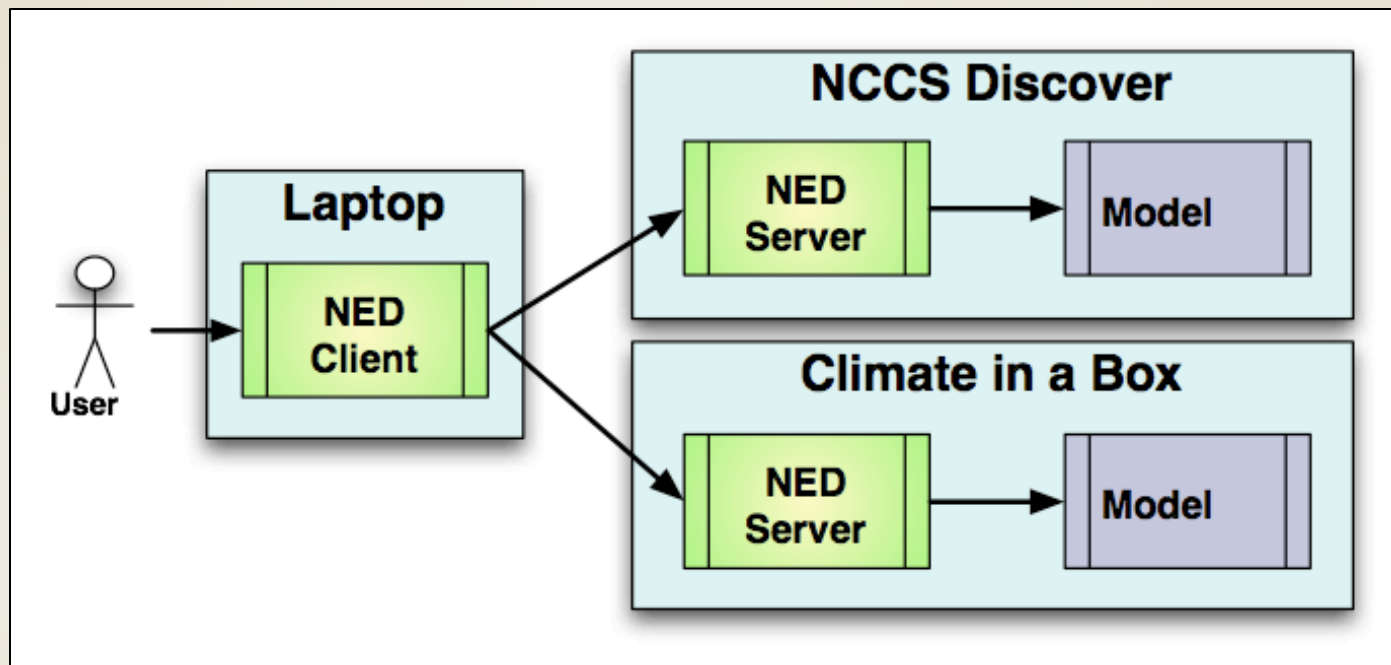






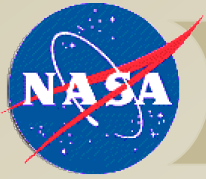
# *NASA Workflow Tool:* *Supports Multiple Sites*

- Use the client and switch to a new location
- Use similar configuration settings to re-run the model at the new location





# *NASA Experiment Designer (NED)*



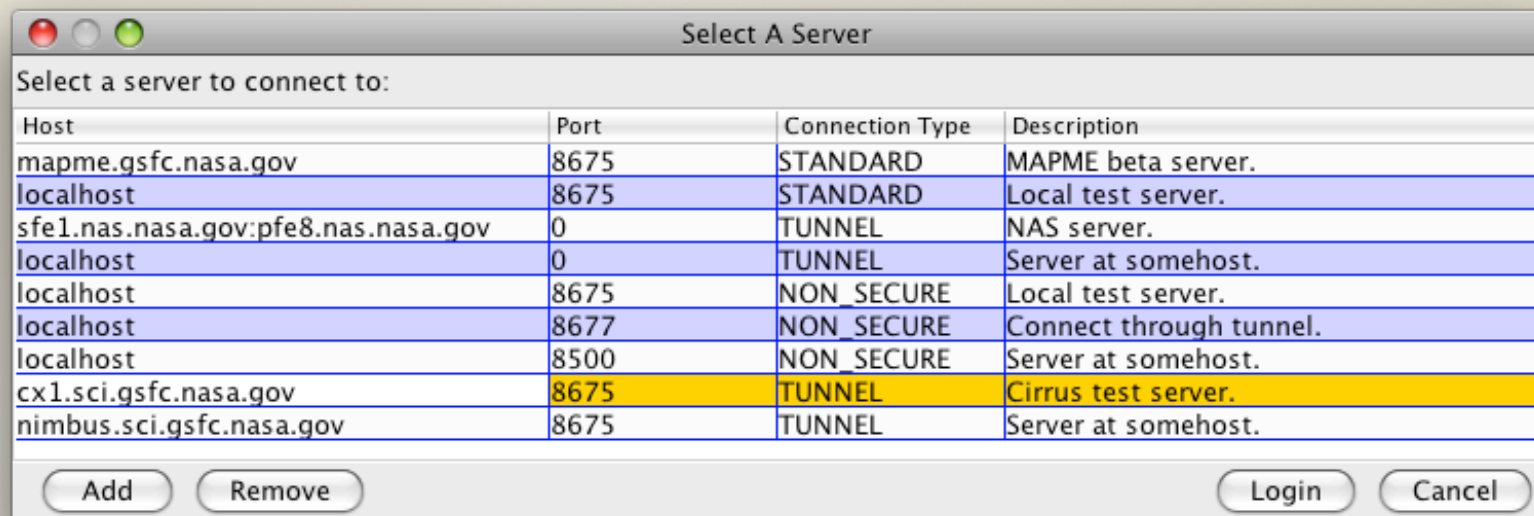
# *NASA Experiment Designer* (NED)

- Consists of a client GUI software and server that runs the workflows underneath.
- Allow user to quickly and easily set up and run an existing workflow
- NED GUI application provides a simple user interface for:
  - Configuring and submitting the workflow
  - Monitoring tasks and task logs as they execute
  - Controlling the tasks in the workflow



# NED Server

- The NED client software connects to a NED server on the host machines
  - NED provides a GUI to manage logins
  - Select a site to connect to, such as “mapme”
  - Must authenticate at the particular site





# *Using the NED Client:*

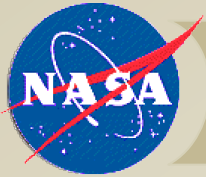
## *Walk-through*

- Starting the NED client
- Window layout
- Loading a configuration (local and remote)
- Making changes to a workflow
- Saving a workflow configuration
- Run modes: As Generic or Yourself
- Submitting a workflow
- Monitoring and controlling the workflow



# *Running a GEOS-5 Workflow*





# *GEOS5 Workflow*

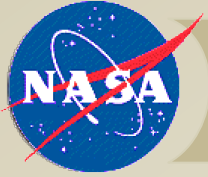
- Supports running on both Discover and CIB systems
- Written in Python
- Supports GEOS5 versions on Discover: MERRA/Eros 25, Fortuna 1.x and 2.0
  - Fortuna 2.1 forthcoming
  - Supports checkout-and-build from GEOS5 repository
- Workflow includes basic restarts for a coarse model run



# *GEOS5 Workflow:*

## *Hands-on Demo*

- Loading the latest workflow configuration
- GEOS5 settings, History Editor
- Connect/run the model from CIB
  - Accessing the results
- Connect/run the model from Discover
  - Accessing the results
- Visualization
  - Using WMS to view results
  - Forthcoming NED feature: view generated images



# *Creating a Workflow*

## *Part I: Overview*



# Workflow Tool Components

NED Client v1.946 User: smfreema Mode: USER

File Search Tools Run Views Help

**Configuration Tree**

- Default Workflow
  - Case\_09\_WPS\_Namelist
    - share
      - wrf\_core
      - max\_dom
      - start\_date
      - end\_date
      - interval\_seconds
      - io\_form\_geogrid
    - geogrid
      - parent\_id
      - parent\_grid\_ratio
      - l\_parent\_start

**Configuration**

Name	Description	Values
Case_Number	Version of WRF being used.	09
User_Name		phayes1
Machine_Name	Machine that WRF will run on.	discover
WPS_Directory	Location of the WRF WPS directory	/discover/nobackup/phayes1/nu-wrf/WPS
WRF_run_Directory	Location of the WRF WRFV3 run directory	/discover/nobackup/phayes1/nu-wrf/WRFV3/run
Total_nodes	Total number of Nodes to use when running wrf.exe	8
Ppn	Processor per node	8
Processor	Which processor to use	harp
Group_List	Group name to use	k3002
Wall_Clock	Wall Clock time for simulation	12:00:00
Intel_Module	Default description.	comp/intel-10.1.021
MPI_Module	Default description.	mpi/mpi-3.2.011
MKL_Module	Default description.	lib/mkl-10.0.5.025
Clean_Up	Do you want the script to clean all the produced files from WPS and run directories? (The wrfout files will be moved to output location)	no

**Status**

Status	Affects	Description
--------	---------	-------------

Current File: /Users/smfreema/Desktop/NU\_WRF\_CASE\_STUDY.nar Logged Into: localhost:8675[Local test server.]

NED Client v1.946 User: smfreema Mode: USER

File Search Tools Run Views Help

**Configuration Tree**

- GEOSGCS
  - Installation
    - Prebuilt
      - INSTALL\_DIR
      - USE\_NED\_RESOURCE
    - New installation
      - OPT\_GMAKE
      - OPT\_CVS
      - CVS\_SYS\_NAME
      - CHEM\_RESOURCE
      - CHEM\_RESOURCE
      - USE\_GS\_MODULES
    - MACH
    - SITE
    - TAG
    - GEOS\_Module
    - MODULE\_ENV
    - RST\_DIR
    - LOCAL\_INSTALL\_DIR
  - Basic settings
    - Run properties
      - EXP\_ID
      - EXPDCS
      - MODEL
    - Resolution and layout
      - MODEL\_RES
      - NX
      - NY
      - DT
    - Run length
      - BEG\_DATE
      - END\_DATE

**Configuration geosgcs\_0026**

geosgcs\_0026

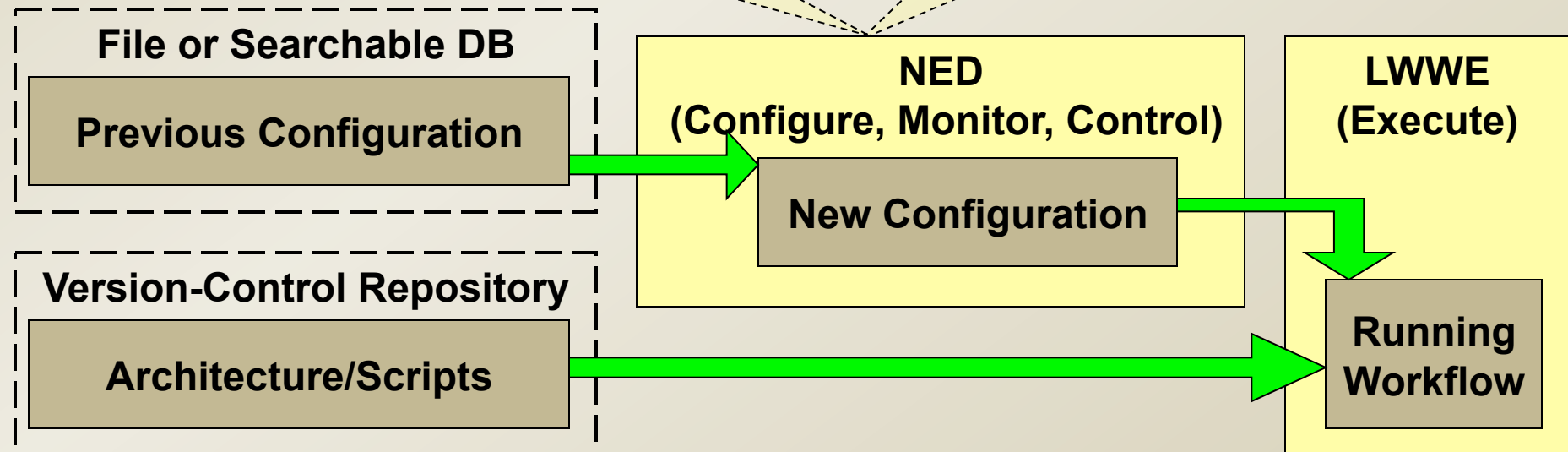
- Prepare Environment Loads NED configuration variables into the workflow.
- Prepare Workflow Setup Rename the top level of experiment directory and then write the bash script to view c
- Setup Workflow Environment Set up the workflow environment
- Setup Model Environment Sets the ocean resolution and sets the forcings and boundary conditions directory.
- Prepare Installation Either copies installation files or installs from CVS. See CVS\_OPT in NED for more informa
- Main Handles the time stepping of the GCM.
  - Create Experiment Directories Creates directories needed for running a GEOS-5 experiment.
  - Create RC Templates Writes the AGCM and CAP resource file templates.
  - Create Live RCS Creates the Live RCS
  - Create History Stream Creates the History Stream directories
  - Link BCS Creates the Link BCS
  - Get GCM Restarts Gets the GCM Restarts
- Run-Loop Handles the run loop tasks [Run 1 of 2]
  - Submit Job Submits the segment ID
  - Wait-In-Queue Waits for the segment ID in queue.
  - Run Model Runs the model after getting the segment ID
  - Rename Checkpoint to Restart Renames Checkpoint to Restart
  - Archive Restarts Archives the restarts
  - Move History Moves the History directory or files

COMPLETED UPDATED QUEUED UNKNOWN SUSPENDED RUNNING RESUMED ABORTED NOT\_SUBMITTED

**Status**

Status	Affects	Description
SUCCESS	GEOSGCS	The workflow was submitted successfully.
INFO	GEOSGCS	The current experiment directory is: /scratch/expWorkDir/geosgcs_0026/
WARN	Installation.New	The variable script ChemistryResourceSource.py executed from

Current File: GEOSGCS Logged Into: mapme.gsfc.nasa.gov:8675[MAPME beta server.]





## *Parts of a Workflow:*

### *Configuration File*

- Typically what is displayed to the user when they are setting up a workflow to run
- Contains:
  - Options that a developer wants to make available to users (e.g. climate model options) and how to display them to the user)
  - Metadata about the workflow itself (e.g. workflow name and where it exists)

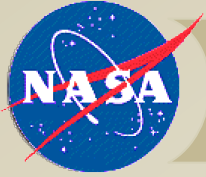


# *Parts of a Workflow:*

## *Architecture File*

- Typically displayed to the user during the execution of the workflow (after submission)
- Contains all the task information necessary to run a workflow (e.g. building a code and running it)

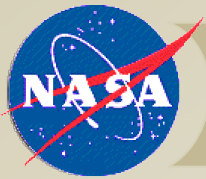




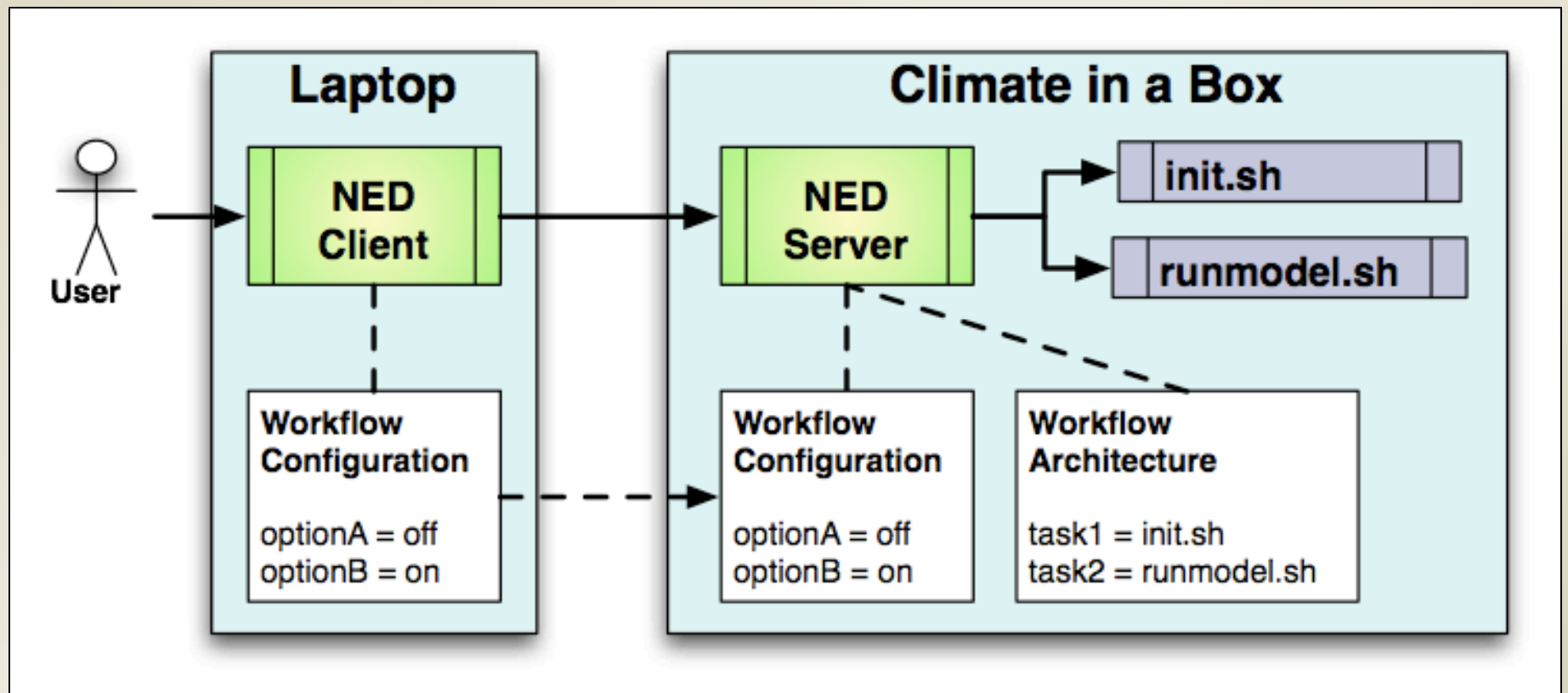
# *Parts of a Workflow:*

## *Other Elements*

- Workflow tasks do activities such as run system commands, shell scripts, or interact with the workflow
  - These are the low level tasks that perform the work for workflow



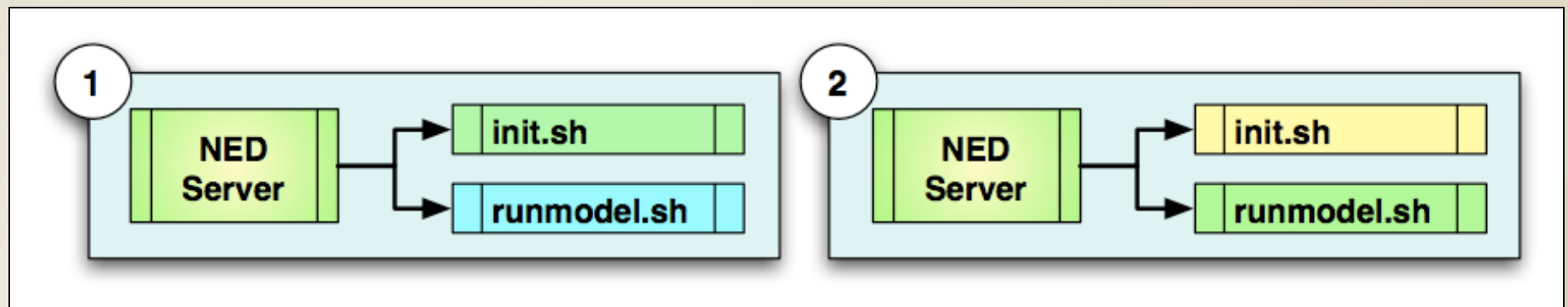
# *Parts of a Workflow: Example*





# *Task Dependencies*

- Workflows operate based on task dependencies
- Task dependencies indicate when each operation occurs
- Example:
  - Task “Run Model” depends on “Init” finishing successfully





# *Designing Workflows*

- Designing good workflows is not simple!
  - Must provide flexible, organized configuration for users
  - Must automate the steps a user normally takes
  - Finding problems can be tricky
  
- The NASA Workflow suite provides tools to help in designing workflows



# *Features for Workflow Developers*

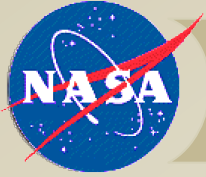
- “Design Mode” feature of the **NASA Experiment Designer (NED)** client software
  - This allows the developer to create a user-friendly configuration
- **Light-Weight Workflow Engine (LWWE) Editor**
  - This allows the developer to specify the various tasks that make up a workflow
- **Workflow templates**
  - This allows users to start from a working workflow and expand it into something useful.



## *NED Designer Interface*

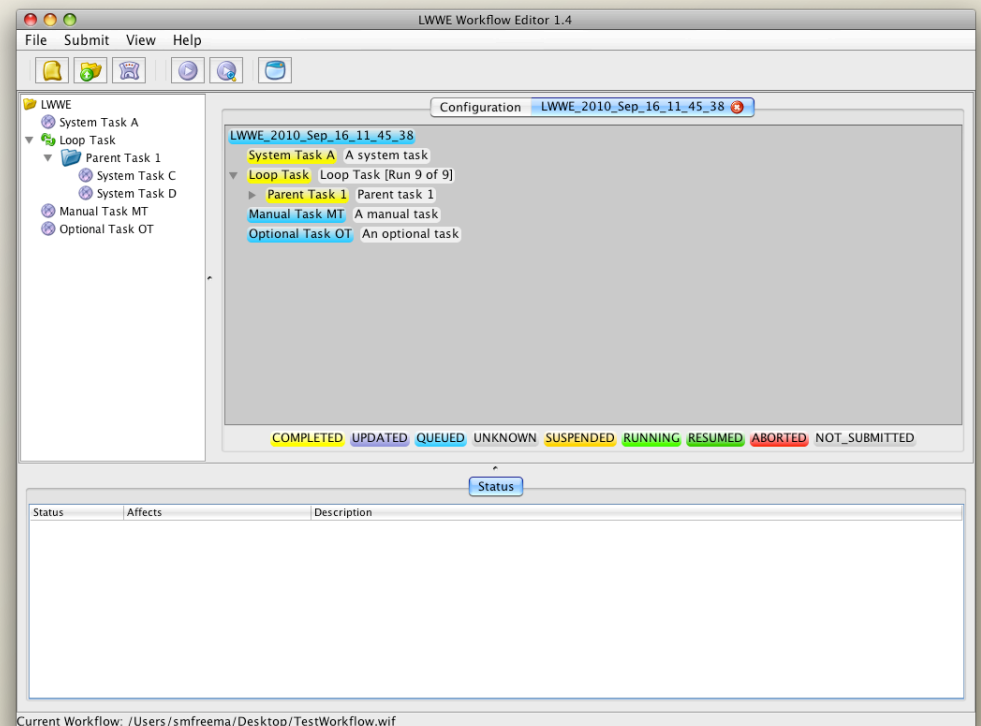
- Available through “File / Design Mode” option in NED
- Makes additional options available:
  - Edit overall workflow properties
  - Add, remove and edit workflow variables and groups
  - Develop configuration scripts for advanced configuration options
    - Example: check for incorrect settings





# *LWWE Editor Tool*

- Allows a workflow developer to quickly construct and validate a workflow architecture
- Test mode allows the developer to test the workflow without running the actual tasks
- Shares a common user interface with NED tool



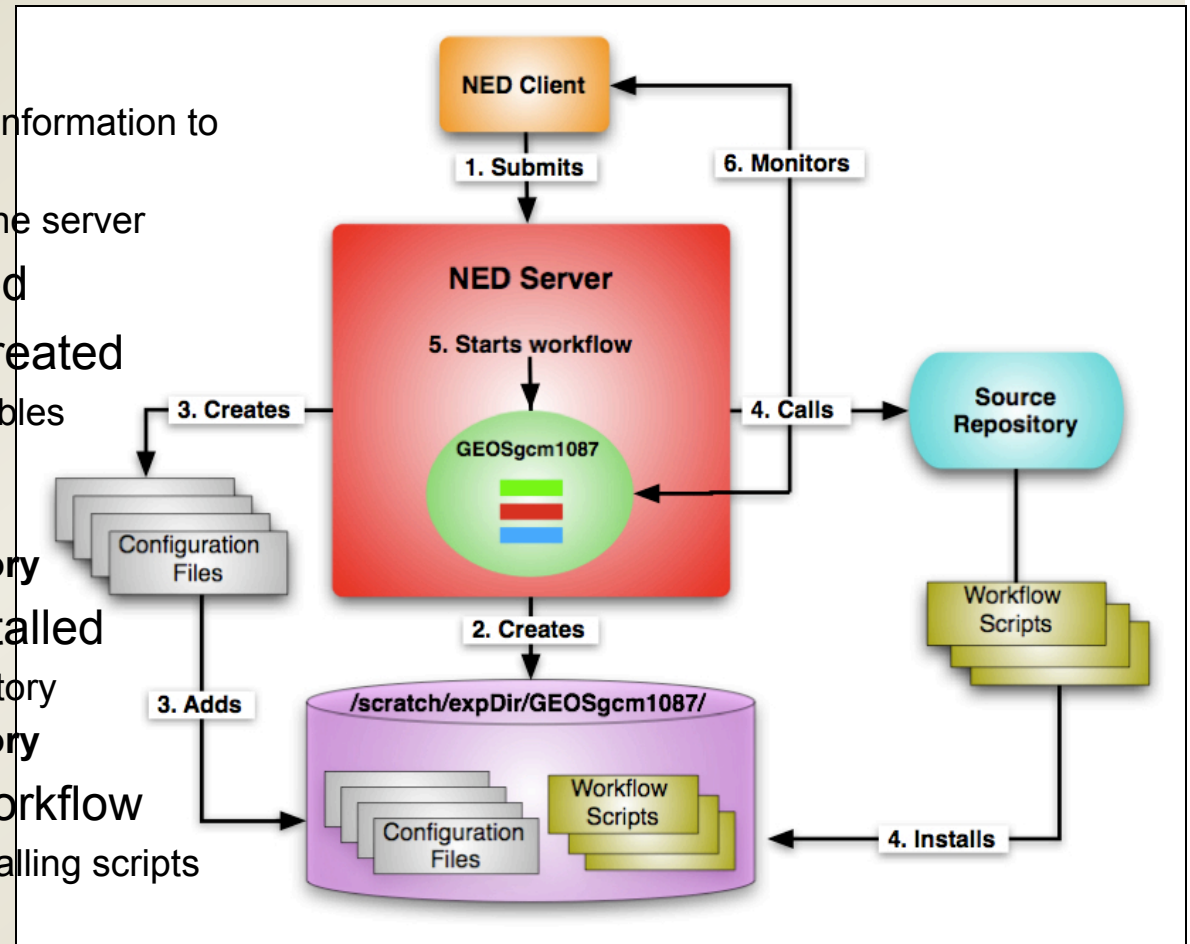


- WRF workflow demo
- Case Study: WRF model workflow
- Quick demo of LWWE editor



# Running a LWWE Workflow in NED

1. User submits workflow
  - NED client sends workflow information to NED server
  - Validation is performed by the server
2. Install directory is created
3. Configuration files are created
  - Define the experiment variables
  - Generated according to the configuration settings
  - Added to the **install directory**
4. Workflow scripts are installed
  - Pulled from a source repository
  - Added to the **install directory**
5. NED server starts the workflow
  - Tasks execute, potentially calling scripts
6. Monitoring occurs
  - Users monitors and controls their workflow





# *Creating a Workflow*

## *Part II: Hands-on Training*



# *Creating a Workflow Demo*

- Demo steps:
  - Access the example workflow
  - Run the example workflow
  - Add a new user setting in the example workflow
  - Add a new task in the example workflow



# Workflow References

- **User Guide and Developer Guide** available on **Modeling Guru**:

The screenshot displays the Modeling Guru website interface. At the top, the NASA logo is on the left, and the title "Modeling Guru" is in large yellow letters. Below the header, a navigation bar includes "Welcome, Guest" and a "Login" link. The main content area is titled "NASA Modeling Guru > MAP Modeling Environment (MAPME) Workflow Tool". Below this, a sub-header reads "MAP Modeling Environment (MAPME) Workflow Tool". A navigation bar with tabs for "Overview", "All Content (37)", "Discussions (17)", "Documents (13)", "Blog", and "Polls" is present. The "Overview" tab is selected, showing a section titled "The MAP ME Workflow Tool" with an "Overview" sub-section. The overview text describes the tool's purpose: "The Workflow tool was created by NASA Goddard Space Flight Center's Software Integration and Visualization Office through its MAP Modeling Environment work. The tool is designed to configure, run, monitor and manage complex model experiments for NASA scientists. Users configure, monitor and manage experiments via an easy-to-use GUI rather than changing configuring scripts and executing system commands. The system takes care of scheduling the individual tasks that make up the workflow to run in the proper order. Such tasks include source check-out, building the executable, execution, post-processing, visualization, etc. Experiments can be customized for many systems and models, and run from one convenient location." Below the text are two screenshots of the tool's interface. The left screenshot shows a "Searchable Experiment DB" with a "Configuration Template" table. The right screenshot shows a "NED (Configuring)" window. Below these screenshots are three yellow boxes labeled "Searchable Experiment DB", "NED (Configuring)", and "XCDP (Monitoring)". On the right side of the page, there is an "Actions" section with a "Notifications" link and a "View feeds" button. Below that is a "Recent Content" section with a "Filter by Categories & Tags" dropdown and a list of recent content items, including "Using the Lightweight Workflow Engine", "Getting Started User Guide", "NED Plug-In Development Guide", "GEOS-5 Workflow Tutorial", "Workflow SDK version 0.10", "Workflow Development Training", "Running the GEOS-5 Workflow with Your Own Restarts", "Workflow Tool Brown Bag Presentation", and "GEOS-5 Workflow Input Files".

<https://modelingguru.nasa.gov/clearspace/community/mapmewkflow>